基于区块链技术的新型分层数字图书馆体系架构*

高胜 朱建明
中央财经大学信息学院 北京 100081

摘要：[**目的/意义**]当前数字图书馆系统数字资源可信性、互操作性及可扩展性等成为阻碍数字图书馆发展的主要问题。[**方法/过程**]基于区块链技术提出一种新型分层数字图书馆体系架构，并详细介绍各层的关键功能组件和核心服务模块，最后与已有相关数字图书馆体系架构进行综合比较。[**结果/结论**]理论分析表明，所提架构利用区块链技术的分布式账本结构、安全可靠、透明可信、时序不可篡改等特征能有效解决数字图书馆数字资源可信性，增强系统架构的互操作性及可扩展性，实现异构数字资源的安全流通。

关键词：数字图书馆 区块链技术 资源可信性 安全性
分类号：G250.76
DOI：10.13266/j.issn.0252-3116.2018.24.008

1 引言

当今，数字化、信息化、智能化等成为社会演进方向。2017年，习总书记在中共中央政治局第二次集体学习指出："推动实施国家大数据战略，加快完善数字基础设施，推进数据资源整合和开放共享，保障数据安全，加快建设数字中国"。数字图书馆作为典型的数字化信息系统，融合了图书馆学、信息系统、数据库管理、多媒体、人机交互等技术[1]。M. A. Gonçalves 等[2]应用5S模型，即数据流(streams)、组织结构(structures)、操作空间(spaces)、脚本(scenarios)及社会团体(societies)，形式化定义数字图书馆。L. Candela 等[3]提出的DELOS参考模型从内容、用户、功能、质量、策略、架构等方面刻画数字图书馆。总之，数字图书馆借鉴图书馆组织模式来管理多种类型的数字化资源，如文档、图片、音频、视频等，利用计算机、网络通讯等信息技术突破时空限制，实现数字资源收集、流转、检索与共享，为用户提供虚拟、高效和便捷的数字化服务。

为满足海量异构信息存储、检索及共享等需求，数字图书馆体系架构是需要解决的基础性问题，其决定了数据图书馆底层基础设施、存储结构、通信协议及服务功能等。C. B. Arms 等[4]提出数字图书馆体系由资源库、句柄系统、检索系统及用户接口组成；张晓林等[5]从数字内容创建、数字对象描述、资源组织描述、数字资源系统服务、数字资源长期保护等方面构建数字信息资源建设的标准描述体系；李爽[6]总结了数字图书馆体系架构的设计原则，即互操作性、可扩展性、规范性、层次化和统一性，其中互操作性和可扩展性是两个需要关注的重点原则；奉国和[7]总结了早期典型的数字图书馆体系架构，包括客户机/服务器两层架构、浏览器/Web服务器/数据库服务器三层架构、客户机/服务器三角形架构、分布式体系架构等。然而，这些体系架构都是构建在传统基础设施上，馆间互操作性、可扩展性、安全性和服务成本等成为主要的瓶颈。随着云计算、服务计算等新兴技术的发展，数字图书馆体系架构得到进一步的发展。现有基于云计算或服务计算数字图书馆体系架构研究中，一部分通过将异构分布式的图书馆资源集成统一平台对外提供服务，是一种典型的中心化架构[8-10]，存在服务效率低、管理成本和风险较大等问题；另一部分研究通过P2P技术构建分布式数字图书馆[11-12]，然而数字资源的可信性依赖于数字资源提供方的信用背书，部分依赖于云基础设施构建的数字图书馆难以防范数字资源伪造、篡改等恶意行为，保障数字资源的可信性。

据2017年S. Higgins报道[13]，美国政府拨款10万美元用以研究区块链在公共图书馆系统中的应用。

* 本文系北京市社会科学基金项目"面向云计算环境北京数字图书馆信息资源安全共享保障体系研究"(项目编号:16XCC023)研究成果之一。
作者简介：高胜(ORCID:0000-0001-8118-411X)，副教授，博士，硕士生导师，E-mail:sgao@cufe.edu.cn；朱建明(ORCID:0000-0003-4923-9939)，教授，博士，博士生导师。
收稿日期:2018-05-08 修回日期:2018-09-04 本文起止页码:57-64 本文责任编辑:徐健

然而,当前基于区块链构建数字图书馆体系非常少。2018年,曾子明和秦思琪[14]利用区块链技术构建了首个面向智慧图书馆移动视觉搜索的去中心化资源管理架构。然而,该架构并没有给出适应数字图书馆具体的区块结构以及服务过程,并且对异构资源互操作性、服务功能扩展性等方面考虑不足。针对上述问题,笔者基于区块链技术构建一种新型分层数字图书馆体系架构。一方面,利用区块链的分布式结构、安全可靠、透明可信、时序不可篡改等特点,克服数字图书馆数据集中管理存在的单点失效、访问瓶颈、数据可信等问题,保障数字图书馆提供数字资源的可信性;另一方面通过自定义统一区块存储结构、密码算法及通信协议等极大增强了异构资源的互操作性及安全性,同时利用区块链的动态性和灵活性等提升数字图书馆系统的可扩展性。

2 相关工作

当前,由于建设条件及技术等因素使得数字图书馆体系架构并不统一。到目前为止,各国已构建了多种不同类型的数字图书馆体系[6-7, 15-16]。1995年,R. Kahn和R. Wilensky[17]较早提出了数字图书馆体系K-W,通过分离独立标识资源库和句柄系统,并利用统一的用户界面实现同一体系下异构数字资源的集中化服务。S. Payette和C. Lagoze等[18]提出灵活可扩展的数字对象及资源库体系架构Fedora,解决数字图书馆内容存储和分发问题。具体地,Fedora通过数字对象模型封装元数据和数据内容以及资源库提供对外服务接口,从而提高异构资源互操作性。针对数字图书馆互操作性问题,文献[16, 19]总结了常见的解决方法,包括标准化,如TCP/IP通信协议、Z39.50检索协议等;中间件技术,如OAI、Dienst等;规范化交互,如代理通信语言、高级语言(SETL、PAISLey)等;移动功能规范,如Lisp、Java applets等。NDSL、NCSTRL等分布式数字图书馆体系架构项目正是采用这些相关技术实现的。2005年P. Knezevic等[20]指出随着面向服务架构(Services-Oriented Architecture, SOA)、P2P网络、网格计算等技术的发展,未来数字图书馆体系将从集成、中心化控制系统向分布式、动态、可配置数字图书馆联盟服务方向演化。李广建等[21]总结了基于SOA、P2P和网格技术构建的数字图书馆体系架构。文献[22-24]研究面向SOA的数字图书馆体系架构。殷红和刘炜[25]指出未来数字图书馆系统中资源发现至关重要。总体而言,现有的架构大都侧重于提高系统互操作性和可扩展性,然而数字资源的可信性完全依赖于资源提供方,增加了数字资源信任建立成本;同时对数字资源安全性考虑不足。

3 区块链技术基础

近年来,作为比特币的底层支撑技术区块链受到国内外广泛关注,其应用已经从区块链1.0数字货币、区块链2.0数字资产等延伸至数字社会等领域[26-28]。笔者将区块链引入数字图书馆领域,构建一种新型的分布式数字图书馆体系架构。

3.1 区块链概念

目前对区块链并没有形成统一的定义。2015年,《经济学人》将区块链描述为创造信任的机器。2016年,工信部发布的白皮书从数据结构和数据处理两个角度定义区块链[29]。具体地,从数据结构角度来看,区块链是一种时序块链式结构,通过密码学保证分布式账本的不可篡改和不可伪造;从数据处理角度来看,区块链是一种分布式基础架构和范式,利用块链式结构存储数据、利用共识算法来生成和更新数据、利用密码学来保证数据安全、利用智能合约来操作数据。总体而言,区块链是一种分布式共享总账,融合了分布式账本技术、密码学、共识算法、智能合约等核心技术,保障在不需要可信第三方信用背书条件下,实现数据的不易篡改、不易伪造、可追溯和审计等特性。

3.2 区块链结构

这里简单介绍区块链数据结构和存储结构。区块链是一种时序块链式数据结构,见图1。其中,每个区块由区块头和区块体构成。不同区块链平台(如比特币区块链、以太坊区块链等)区块结构是不一样的。中国区块链技术和产业发展论坛上发布的《区块链数据格式规范》标准草案中给出了区块、事务、实体等数据结构[30]。通常,区块头存储上一个区块的哈希值、Merkle根哈希值、随机数等,区块体存储某段时间内历史交易数据信息。当前区块的区块头存储上一区块的双哈希值形成时序链式结构。Merkle根哈希能快速校验区块交易的完整性,使得在不同步所有交易数据情况下可实现对特定交易的验证,即"简单支付验证",从而提高区块链运行效率和可扩展性。区块链存储结构采用键值对形式,从而方便了交易、区块等数据检索操作。存储方式可以文件方式存储在磁盘,也可以以数据库方式存储。当前比特币区块链中数据以文件形式存储,索引数据存在LevelDB中;而以太坊区块链数据和索引数据都存在LevelDB中。

3.3 区块链关键技术

区块链是分布式账本、密码学、共识算法和智能合约等多种技术综合[27]。

(1)分布式账本。不同类型区块链采用的账本模型是不一样的。目前,常见的账本模型包括基于交易的账本模型(如比特币、量子币等)和基于账户的账本模型(如以太坊、超级账本等)。

(2)密码技术。区块链主要采用数字签名技术来保证交易的完整性和不可抵赖性等。每个用户拥有一对公私钥对,其中公钥用来产生交易地址,私钥用来对签名交易数据。

(3)共识算法。根据参与主体准入机制和共识算法的不同,区块链可分为公有链、联盟链和私有链,其中联盟链和私有链统称为许可链。不同类型区块链所采用的共识算法是不一样的,如比特币、以太坊等公有链大都采用工作量证明,具有交易性能低、资源开销大等问题;超级账本联盟链采用实用拜占庭容错(Practical Byzantine Fault Tolerance,PBFT)算法[26],可提供较高交易性能和可扩展性。

(4)智能合约。N. Szabo 最先提出智能合约,并指出其是执行合约条款的可计算交易协议[31]。2013 年,V. Buterin[32]将智能合约引入以太坊,提出了内置图灵完备编程语言的去中心化应用平台,将智能合约定义为一个运行在安全环境下的计算机程序,可以直接控制数字资产。为了防止智能合约漏洞影响主机的安全性,一般智能合约都运行在具有隔离特性的沙箱环境。

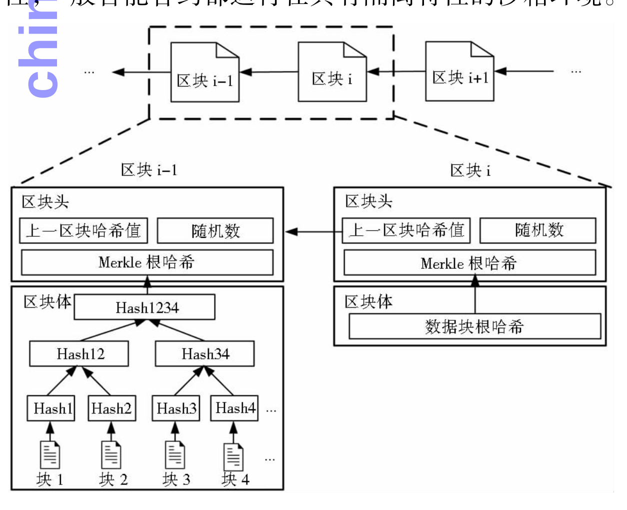


图 1 区块链数据结构

4 基于区块链技术的数字图书馆体系架构设计

区块链的技术特性可有效增强分布式环境下异构数字资源的互操作性和可信性,提高数字图书馆可扩展性和服务质量。基于此,笔者基于区块链技术设计了一种独立自治、分层管理的新型数字图书馆体系架构,如图 2 所示:

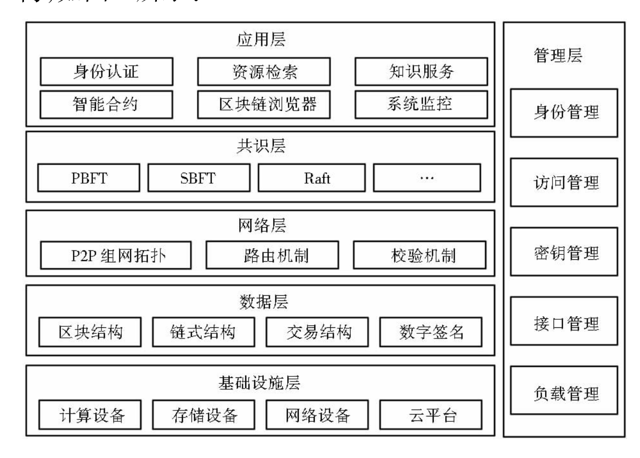


图 2 基于区块链的数字图书馆体系架构

4.1 基础设施层

基础设施层提供构建数字图书馆体系架构的软硬件设施,主要包括计算设备、存储设备、网络设备等。具体地,计算设备是一组分布式服务器节点,提供区块链相关操作的计算环境。存储设备用来存储区块数据、事务数据、链式结构、合约代码、元数据、数字资源及其他数据等,其中数字资源主要是以文本、图像、音频、视频等表现形式的文献数据,如电子期刊/会议论文、学术论文、电子书等[5]。当前常见的存储管理系统有文件系统(如磁盘、移动硬盘等)和数据库系统(如 LevelDB、Oceanbase 等)。网络设备为区块链节点提供网络通信环境,包括网卡、交换机、路由器等,实现区块、交易、查询等数据传输。

目前,数字图书馆的基础设施层可由各单位投入资金购买物理设备实现本地部署。然而,这种方式带来了较高建设和维护成本,占用较多机房空间;同时,系统弹性受限于所购买的物理设备,即物理设备性能过载或冗余容易导致服务质量下降或资源浪费。另一类可通过租赁第三方云平台搭建,常见的云平台有 IBM Bluemix、亚马逊 AWS 等。利用云计算的虚拟化技术将分布式、异构的物理资源(包括计算、存储、网络等设备)抽象为资源池,屏蔽底层物理资源设备的差异性,为上层提供统一、透明、按需付费使用的基础设施即服务(Infrastructure as a Service, IaaS)。通过 IaaS 可有效整合分布式、异构的物理资源,降低用户部署和维护成本,减少机房占用空间,提高系统资源利用率;同时用户根据业务需求自适应购买或释放所需物理资

源,实现系统弹性可扩展或可伸缩。

4.2　数据层

为增强数字图书馆体系架构异构资源互操作性,数据层定义了数字图书馆统一资源存储结构和管理方式,主要包括区块结构、链式结构、交易结构及数字签名等。

当前,不同应用所构建的区块链数据结构是不一样的。薛腾飞等[33]提出面向医疗领域的区块链数据共享模型,并设计了由区块头和区块体组成的医疗数据存储结构,其中区块体存储医疗数据包括数据所有者公钥、元数据和数据摘要。为了增强数字图书馆体系异构资源互操作性,笔者基于数字资源类型和区块数据结构设计了一种面向数字图书馆的统一数据结构,如图 3 所示:

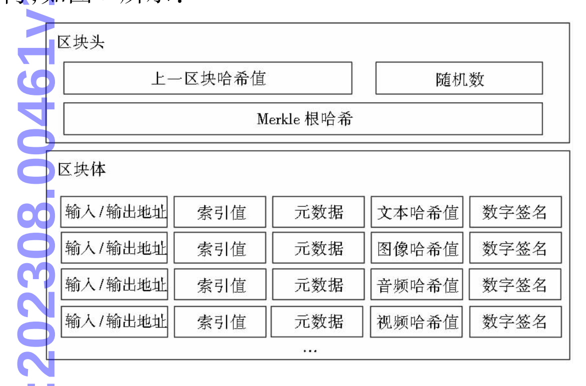


图 3　数字图书馆统一的区块数据结构

所设计的面向数字图书馆的统一数据结构由区块头和区块体组成,存储于区块链网络中的每个节点,其中区块头用以链接区块、验证数字资源的完整性等,即在当前区块头中存储上一区块哈希值实现链式结构;区块体用以记录某段时间内的交易信息,主要包含数字图书馆中不同类型数字资源记录,例如文本、图像、音频、视频等,并生成 Merkle 根哈希写入到区块头。每条数字资源记录由索引值、元数据和数字资源哈希值构成,其中索引值代表数字资源特征,用以提高交易记录检索效率。一般地,可利用信息增益、互信息、交叉熵等函数提取数字资源特征。元数据用来描述数据资源属性,主要包括数据元素、数据格式、数据交互等[34]。笔者采用诸如 MARC、AACR2、Dublin core 等通用标准表示异构数字资源,并以 HTML 或 XML 等标记语言显示,通过 TCP/IP、Z39.50 等标准化数据通信协议交互元数据,实现了对数字图书馆中复杂数字资源统一表示、存储及交互,从而增强分布式异构资源的互操作性;数字资源哈希值一方面作为索引值检索存储在本地文件系统或数据库系统中数字资源,另一方面可用来验证数字资源的完整性。

数字图书馆中所有数字资源操作都是通过区块链交易完成,包括数字资源获取和数字资源检索等。区块链记录了数字资源的交易信息,交易信息被广播给区块链网络节点,每隔一段时间由其打包到区块并通过共识层提供的共识算法生成新区块,然后链接在原来区块链尾部。交易数据结构包含资源提供方地址、接收方地址、索引值、元数据、数字资源哈希值、数字签名等信息。区块链中每个节点拥有一对公私钥对,其中公钥用来生成交易地址,私钥用来签名交易信息。数字签名模块提供 ECDSA 或 RSA 等签名算法完成对每笔交易信息的数字签名,并通过公钥进行验证,以保障交易的完整性和不可否认性。

4.3　网络层

网络层主要提供 P2P 组网、路由机制及验证机制等功能,实现数字图书馆区块和交易等数据传输。例如,比特币和以太坊采用基于 TCP 的 P2P 协议,Hyperledger 采用基于 HTTP/2.0 的 P2P 协议[26]。数字图书馆系统架构采用 P2P 网络组织区块链节点,每个节点既是数字资源提供方,也是使用方。系统中不存在中心化服务节点,从而缓解了集中式服务架构带来的单点失效、性能瓶颈等问题,提高了系统的可扩展性。当前,P2P 组网拓扑可分为结构化 P2P 网络和非结构化 P2P 网络,其中结构化 P2P 网络主要通过分布式哈希表来组织网络节点,如 Chord、Pastry 等;非结构化 P2P 网络包括中心化拓扑(如 Napster)、全分布式非结构化拓扑(如 Gnutella)、混合式拓扑(如 KaZaa)。

路由机制提供了节点发现、路由选择、区块或交易广播等功能。节点发现功能实现新加入节点链接到区块链网络中稳定节点。由于 P2P 节点与地理位置无关,因而可以通过节点发现功能链接若干个节点接入区块链网络,然后由这些节点链接其他节点实现全网互联,从而可实现区块或交易广播功能。校验机制用来验证区块或交易数据的有效性,避免无效数据的广播浪费网络资源。每个节点在收到区块或交易数据后,需要从数据结构、数字签名、共识算法等方面验证区块或交易的有效性。若有效,则将区块链接到原区块链尾部或将交易打包到新区块中;否则就丢弃该数据,阻止其继续传播。

4.4　共识层

数字图书馆中数字资源操作以交易形式发布到链上,由共识层负责执行。共识层通过用户需求和数字图书馆数字资源之间匹配度达成一致性意见产生数字

高胜，朱建明. 基于区块链技术的新型分层数字图书馆体系架构[J]. 图书情报工作,2018,62(24):57－64.

图书馆新区块,解决分布式环境下数字图书馆节点存储数字资源的一致性及可信性问题。当前不同类型的区块链系统所采用的共识算法是不一样的,主要可以分为:基于计算资源类共识和基于投票类共识。由于公有链的自由性和开放性,为了抵御女巫攻击,公有链一般采用基于计算资源类的共识算法,即通过节点计算力替代节点投票来竞争记账权保障节点存储区块的一致性。例如,比特币采用工作量证明机制(Proof of Work, PoW);以太坊当前使用 PoW,后期计划切换到权益证明(Proof of Stake,PoS)等。联盟链提供节点管理、认证、审计、控制等功能,一般采用基于投票类的共识算法,例如,Hyperledger 的实用拜占庭容错 PBFT、Quorum 的 QuorumChain 共识。

当前,数字图书馆已形成多个联盟性组织,如中国数字图书馆联盟(CDLF)、中国高等学校数字图书馆联盟(CALIS)及各地方数字图书馆联盟等。考虑到公有链共识效率较低、计算开销大等问题,笔者基于现有已形成的数字图书馆联盟提出的基于区块链数字图书馆体系架构采用基于投票类的共识算法,如 PBFT、SBFT、Raft 等[26],一方面由于数字图书馆联盟成员节点数量有限且固定,成员间网络连接较为稳定,因而区块产生和验证过程只需要在联盟成员节点间进行,而并不需要全网所有节点参与,因此仅通过联盟成员之间进行消息投票达到交易即确认,而非任意节点通过购买大量计算资源来实现概率性共识,从而能在保证成本的前提下,减少区块产生时间,保障数字资源可信性的同时极大地提高交易效率。另一方面数字图书馆联盟可对区块链数据设置访问权限,一定程度上保障了数字资源安全。此外,由于数字图书馆新区块产生是由预先选好的节点决定,其他节点只参与验证过程,因而数字图书馆联盟可以不需要经济激励机制。

4.5 应用层

应用层为用户提供数字图书馆分布式应用服务,包括区块链服务及用户服务。

区块链服务主要包括智能合约、区块链浏览器及系统监控等。①智能合约服务通过将数字图书馆中业务逻辑数字化、程序化、自动化等,一方面简化了诸如文献传递、资源导航、信息推送等数字图书馆服务流程,降低中间环节数字资源盗窃和欺诈的风险,提高数字资源可信性及服务效率,减少信用背书及服务成本。例如,通过智能合约可实现无人值守和干预的数字图书馆文献传递,从而提高数字资源流转自动化和智能化。另一方面提供灵活、可编程的数字图书馆功能,用户可通过 Solidity、C + + 、Java 等编程语言定制满足需求的智能合约,并通过接口部署到区块链上,保障数字资源的可信性,同时实现数字图书馆的个性化服务。②区块链浏览器以 Web 形式提供区块及交易信息查询及浏览功能。用户可根据区块哈希值、交易哈希值、区块高度、交易地址等信息查询。例如,用户可通过区块哈希值或区块高度等查询到区块数据,包括区块头和区块体,具体数据结构见数据层;通过交易哈希值查询某笔交易具体信息,包括交易双方地址、索引值、元数据、数字资源哈希值、数字签名、确认数等信息。浏览功能主要查看最近的数字资源交易记录、每日交易量、交易吞吐量、交易速率等信息。③系统监控主要以可视化形式实时显示系统资源耗费及负载情况,包括 CPU 使用率、进程数、内存利用率等信息。

用户服务主要包括身份认证、资源检索和知识服务等。①身份认证提供用户注册、登录、注销等功能。只有通过身份认证的合法用户才被允许进入基于区块链的数字图书馆联盟链网络,并根据所授权的权限进行数字图书馆联盟链上的相关操作。当前数字图书馆联盟链认证可采用口令认证、公钥证书认证、生物特征认证等方式。②资源检索是通过区块链交易实现,是数字图书馆提供的主要服务。笔者首先采用 TF-IDF 方法[35]提取用户查询内容的特征值,并将请求以区块链交易形式提交给数字图书馆联盟链。然后,联盟链节点通过计算区块体中索引值与特征值之间的相似关系,通过阈值法或 top-n 策略找到满足要求的交易记录,即相似关系高于所定义的阈值或者相似关系排在前 n 即满足要求。最后,根据交易记录中的数字资源哈希值查找本地文件系统或数据库得到所需数字资源并返回给用户。③知识服务提供产权保护、数据审计与分析等数字图书馆服务。知识产权可通过数字签名实现,数据审计与分析可利用大数据技术完成。

4.6 管理层

管理层负责维护基于区块链的数字图书馆体系架构,包括身份管理、访问管理、密钥管理、接口管理、负载管理等。①身份管理一方面负责数字图书馆联盟成员数字身份的增加、更新、删除等操作,实现数字图书馆联盟链中节点的动态加入和退出;其中数字身份是用以证明自身合法性的信息或属性,具有唯一性、稳定性等特点,如身份标识(身份证、邮箱等)、生物特征(指纹、虹膜等)。另一方面负责身份保密和鉴别,解决身份泄露和伪造等攻击,提供管理身份的安全性和隐私性[36]。当前身份保密一般采用加密技术、匿名技

术、差分隐私等实现;身份鉴别可采用不同类型身份认证技术实现。②访问管理一方面规定了数字图书馆联盟链中合法节点对数字资源的访问权限和策略。访问权限可由访问控制矩阵、访问控制列表、能力关系表等建模。基本的访问控制策略有自主访问控制、强制访问控制、角色访问控制等。考虑到数字图书馆中数字资源安全需求的差异性,笔者引入等级保护技术实现对数字图书馆数字资源的分级保护。③密钥管理负责管理数字图书馆联盟链中每个节点的公私钥对,可采用分层钱包实现。④接口管理一方面为分层数字图书馆体系提供层与层之间调用的接口函数;另一方面提供了外部系统访问智能合约服务、数字图书馆联盟链平台及数据接口。⑤负载管理主要根据当前各个节点资源使用率进行任务调度,优化区块链节点性能,为共识达成、智能合约执行等过程的高效性提供资源保障,提高数字图书馆资源存储和检索性能。

5 体系架构比较与分析

本节将从体系架构、互操作性、可扩展性、安全性、资源可信性等方面对比笔者提出的架构与相关架构的关系,如表 1 所示:

当前,数字图书馆体系架构主要包括基于 SOA 体系[22-24]、云计算体系[8-10]、点对点体系[11-12]等。SOA 数字图书馆体系架构[22-24]和云计算数字图书馆体系[8-10]都是将数字资源整合在一起对外提供统一服务,具有较好的互操作性和可扩展性。然而这些逻辑上都是一种中心化架构,存在服务性能瓶颈、单点失效、安全风险较大等问题;点对点数字图书馆架构[11-12]通过分布式方式管理数字化资源并对外提供服务,具有较好的互操作性和可扩展性。然而,这三类架构中数字资源的可信性完全依赖于数字资源提供者的信用背书,增加了数字资源信任建立成本;同时没有考虑数字图书馆中数字资源的安全性。例如,云计算数字图书馆构建在云端使得数字资源所有权和管理权相分离,从而极大威胁数字资源的安全性。文献[14]利用区块链技术构建了首个面向智慧图书馆移动视觉搜索的去中心化资源管理架构。然而,该架构并没有给出适应数字图书馆具体的区块结构以及服务过程,并且对异构资源互操作性、服务功能扩展性等方面考虑不足。相比现有这些数字图书馆体系架构,在数字资源可信性方面,本架构利用区块链的分布式账本、密码学、共识算法等关键技术保障了所构建的数字图书馆体系架构能在不依赖于第三方机构信任背书情况下保证数字资源的可信性。在互操作性方面,本架构通过自定义面向数字图书馆统一数据结构,包括区块存储结构、交易结构等,屏蔽了不同数字图书馆中数字资源存储形式的差异性,以及密码算法和通信协议,从而有效增强了不同数字图书馆中数字资源的互操作性。在可扩展性方面,本架构利用云计算技术提供 IaaS 实现了系统弹性可扩展;同时采用联盟链的部署形式增强了数字图书馆节点的动态性和可扩展性,加快了共识达成效率;此外根据节点资源使用率进行有效负载管理,优化区块链节点性能,保障了数字资源可信性的同时提高了数字资源的获取效率。

然而,由于区块链技术本身仍然不完善,限于篇幅受限,所提出的基于区块链的数字图书馆体系架构未来将研究:①所构建的数字图书馆联盟链是动态增长、不可篡改的,随着交易数据的增加,区块体积膨胀及性能问题会越来越突出。现有研究提出了一些解决方案,例如利用隔离见证扩大区块容量,优化数据结构;通过分片技术、闪电或雷电网络等进一步增加系统性能。②交易数据由联盟图书馆共同维护,然而不同数字图书馆对数字资源的隐私需求不同,为保障数字资源隐私性,对节点访问、查询等进行有效管理仍然需要进一步研究。③结合所提出的基于区块链的数字图书馆体系架构,开发与现有数字图书馆相兼容的应用系统,推动落地实践是下一步需要开展工作。

表 1 数字图书馆体系架构比较

对比项 / 相关工作	体系架构	互操作性	可扩展性	安全性	资源可信性
文献[22-24]	SOA 数字图书馆架构	√	√	×	×
文献[8-10]	云计算数字图书馆架构	√	√	×	×
文献[11-12]	点对点数字图书馆架构	√	√	×	×
文献[14]	去中心移动搜索架构	×	×	√	√
本架构	区块链数字图书馆架构	√	√	√	√

6 总结

针对当前数字图书馆体系架构难以防范数字资源伪造、篡改等恶意行为,笔者提出了一种基于区块链技术的数字图书馆体系架构,解决了数字资源可信性问题,降低了数字资源信任构建成本。相比于现有数字图书馆体系架构,该架构通过自定义统一区块存储结构、标准化密码算法及通信协议等增强了异构资源的互操作性及安全性,同时系统基础设施、区块链部署形式及服务负载优化等提升了数字图书馆系统可扩展性。未来将围绕数字图书馆联盟链区块体积膨胀、性能优化、数字资源隐私保护等方面研究,进一步推动所

提出的数字图书馆体系架构落地实践。

参考文献：

[1] LEONARDO C, CASTELLI D, PAGANO P, et al. Setting the foundations of digital libraries: the DELOS manifesto[J]. D-Lib magazine, 2007, 13(3/4):doi:10.1045/march2007-castelli.

[2] GONCALVES M A, FOX E A, WATSON L T, et al. Streams, structures, spaces, scenarios, societies (5s): a formal model for digital libraries[J]. ACM transactions on information systems, 2004, 22(2):270－312.

[3] CANDELA L, CASTELLI D, FERRO N, et al. The DELOS digital library reference model. Foundations for digital libraries(Version 0.98) [M]. Pisa: ISTI-CNR at Gruppo ALI, 2008:1－215.

[4] ARMS C B, OVERLY E A. An architecture for information in digital libraries[J/OL]. [2018－09－03]. http://www.dlib.org/dlib/february97/cnri/02arms1.html.

[5] 张晓林，曾蕾，李广建，等. 数字图书馆建设的标准与规范[J]. 中国图书馆学报，2002，28(6):7－16.

[6] 李爽. 数字图书馆体系结构的设计原则与系统构成[J]. 图书情报知识，2004，2(1):52－54.

[7] 奉国和. 数字图书馆体系结构研究[J]. 情报杂志，2009，28(7):145－149.

[8] 张红丽. 基于云计算平台的分布式数字图书馆框架模型研究[J]. 情报科学，2013，31(3):40－67.

[9] 王文清，陈凌. CALIS 数字图书馆云服务平台模型[J]. 大学图书馆学报，2009，27(4):13－18,32.

[10] 喻昕，王敬一. 基于云计算技术的数字图书馆云服务平台架构研究[J]. 情报科学，2011，29(7):1049－1053.

[11] 刘成山，王洁良，秦春秀. 一种基于对等云的数字图书馆架构[J]. 情报理论与实践，2013，36(11):114－117.

[12] PODNAR I, LUU T, RAJMAN M, et al. A peer-to-peer architecture for information retrieval across digital library collections [C]// GONZALO J, THANOS C, VERDEJO M F, et al. ECDL'06. Berlin: Springer, 2006: 14－25.

[13] HIGGINS S. LibraryChain? US government grants 100k for new blockchain research[EB/OL]. [2017－09－01]. https://www.coindesk.com/librarychain-us-government-grants-100k-new-blockchain-research/.

[14] 曾子明，秦思琪. 去中心化的智慧图书馆移动视觉搜索管理体系[J]. 情报科学，2018，36(1):11－15.

[15] 贺延辉. 俄罗斯国家数字图书馆的建设与发展[J]. 中国图书馆学报，2014，40(5):35－51.

[16] 张晓林. 分布式数字图书馆机制[J]. 情报学报，2002，21(1):63－70.

[17] KAHN R, WILENSKY R. A framework for distributed digital object services[J]. International journal on digital libraries, 2006, 6(2):115－123.

[18] PAYETTE S, LAGOZE C. Flexible and extensible digital object and repository architecture (FEDORA) [C]// NIKOLAOU C, STEPHANIDIS C. ECDL'98. Berlin: Springer, 1998: 41－59.

[19] PAEPCKE A, CHANG C, GARCIA-MOLINA H, et al. Interoperability for digital libraries worldwide[J]. Communications of the ACM, 1998, 41(4):33－42.

[20] KNEZEVIC P, MEHTA B, NIEDEREE C, et al. Supporting information access in next generation digital library architectures[C]// TURKER C, AGOSTI M, SCHEK H J. Peer-to-Peer, grid, and service-orientation in digital library architectures. Berlin: Springer, 2005:207－222.

[21] 李广建. 下一代数字图书馆系统体系架构的发展动向[J]. 高校图书馆工作，2006，26(113):1－5.

[22] 侯福丽. 基于 SOA 的数字图书馆平台体系架构[J]. 情报科学，2007，25(6):903－908.

[23] DONG L, XING C X, LIN J, et al. Personalized digital library framework based on service oriented architecture [C] // BUCHANAN G, MASOODIAN M, CUNNINGHAM S J. ICADL'08. Berlin: Springer, 2008: 351－354.

[24] 易菲，龙朝阳. 基于 SOA 的数字图书馆联盟信息服务平台构建[J]. 图书馆工作与研究，2011(6):4－7.

[25] 殷红，刘炜. 新一代图书馆服务系统:功能评价与愿景展望[J]. 中国图书馆学报，2013，39(5):26－33.

[26] 邵奇峰，金澈清，张召，等. 区块链技术:架构及进展[J]. 计算机学报，2018,41(5):969－988.

[27] DINH T T A, LIU R, ZHANG M, et al. Untangling blockchain: a data processing view of blockchain systems[J]. IEEE transactions on knowledge and data engineering, 2018, 30(7): 1366－1385.

[28] 袁勇，王飞跃. 区块链技术发展现状与展望[J]. 自动化学报，2016，42(4):481－494.

[29] 工业和信息化部. 中国区块链技术和应用发展白皮书[R/OL]. [2017－04－08]. http://www.cbdforum.cn/bcweb/index/article/rsr-6.html.

[30] 中国区块链技术和产业发展论坛. 区块链数据格式规范[R/OL]. [2017－12－27]. http://www.cesi.cn/cesi/201712/3465.html.

[31] SZABO N. Smart contracts[EB/OL]. [2017－01－01]. http://www.fon.hum.uva.nl/rob/Courses/InformationInSpeech/CDROM/Literature/LOTwinterschool2006/szabo.best.vwh.net/smart.contracts.html.

[32] BUTERIN V. A next-generation smart contract and decentralized application platform[R/OL]. [2017－04－08]. https://github.com/ethereum/wiki/wiki/White-Paper.

[33] 薛腾飞，傅群超，王枞，等. 基于区块链的医疗数据共享模型研究[J]. 自动化学报，2017(9):1555－1562.

[34] GROUP D A. Descriptive metadata guidelines for RLG cultural materials [EB/OL]. [2017－04－15]. https://www.oclc.org/content/dam/research/activities/culturalmaterials/RLG_desc_metadata.pdf.

[35] SALTON G, BUCKLEY C. Term-weighting approaches in automatic text retrieval[J]. Information processing & management, 1988, 24(5):513－523.

[36] 吴宗大, 谢坚, 郑城仁, 等. 数字图书馆用户的行为偏好隐私保护框架[J]. 中国图书馆学报, 2018, 44(2):72－85.

作者贡献说明:

高胜:负责论文的整体构架、撰写及修改;

朱建明:负责论文修改。

A Novel Hierarchical Digital Library System Architecture Based on Blockchain Technology

Gao Sheng　Zhu Jianming
School of Information, Central University of Finance and Economics, Beijing 100081

Abstract: [**Purpose/significance**] At present, the digital resource problems of trust, interoperability and scalability in digital library system have become the major obstacle for the development of digital library. [**Method/process**] Based on the blockchain technology, we proposed a novel hierarchical digital library system architecture. Then, we described the key functional components and core service modules of each layer in detail. Finally, we compared our proposed digital library system architecture with the others. [**Result/conclusion**] Theoretical analysis proves that our proposed digital library system architecture exploits the features of blockchain, such as distributed ledger architecture, security and reliable, transparent and credible, time-series and tamper-resistant, can solve the trust problem of digital resource, enhance the interoperability and scalability and achieve the security flow of heterogeneous digital resource.

Keywords: digital library　blockchain technology　trust　security